

# Data Preparation

*Alex J. Benecke*

*June 15, 2017*

```
LMB <- read.csv("Data/Raw-Data/2016_largemouth-bass_raw.csv") %>%
  dplyr::select(Month:Site, AgeCap:Sex) %>%
  arrange(FID)

str(LMB) ### n = 132

## 'data.frame': 132 obs. of 20 variables:
## $ Month : int 6 6 6 6 6 6 6 6 6 6 ...
## $ Day : int 27 27 27 27 29 29 29 29 29 29 ...
## $ Year : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ FID : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Site : int 11 11 11 12 18 18 18 18 18 18 ...
## $ AgeCap: int 4 4 4 6 4 NA 1 2 4 8 ...
## $ Anu1 : num 0.433 0.567 0.462 0.462 0.405 ...
## $ Anu2 : num 0.69 0.718 0.636 0.645 0.559 ...
## $ Anu3 : num 0.803 0.849 0.751 0.784 0.72 ...
## $ Anu4 : num 0.927 0.958 0.816 0.878 0.799 ...
## $ Anu5 : num 0.94 0.988 0.836 0.998 0.832 ...
## $ Anu6 : num NA NA NA 1.04 NA ...
## $ Anu7 : num NA NA NA 1.08 NA ...
## $ Anu8 : num NA NA NA NA NA ...
## $ Anu9 : num NA NA NA NA NA ...
## $ RadCap: num 0.94 0.988 0.836 1.079 0.832 ...
## $ LenCap: int 347 292 348 374 375 355 195 289 388 423 ...
## $ WTg : int 658 415 557 669 716 719 118 479 986 1258 ...
## $ SEXCON: int 8 3 3 8 3 8 8 3 3 8 ...
## $ Sex : int 2 1 1 2 1 2 2 1 1 2 ...

tmp <- LMB %>% filter(!is.na(AgeCap))
length(tmp$AgeCap) ### Aged fish = 127, not aged = 5

## [1] 127

mean(tmp$AgeCap)

## [1] 2.669291

median(tmp$AgeCap)

## [1] 2

range(tmp$AgeCap)

## [1] 1 8

table(tmp$Sex)

##
## 0 1 2
## 1 56 70
```

```

### Making factors factors
LMB$FID <- factor(LMB$FID)
LMB$Site <- factor(LMB$Site)
LMB$SEXCON <- factor(LMB$SEXCON)
LMB$Sex <- factor(LMB$Sex)
str(LMB)

## 'data.frame': 132 obs. of 20 variables:
## $ Month : int 6 6 6 6 6 6 6 6 6 6 ...
## $ Day : int 27 27 27 27 29 29 29 29 29 29 ...
## $ Year : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ FID : Factor w/ 132 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Site : Factor w/ 11 levels "2","4","6","8",...: 6 6 6 7 10 10 10 10 10 10 ...
## $ AgeCap: int 4 4 4 6 4 NA 1 2 4 8 ...
## $ Anu1 : num 0.433 0.567 0.462 0.462 0.405 ...
## $ Anu2 : num 0.69 0.718 0.636 0.645 0.559 ...
## $ Anu3 : num 0.803 0.849 0.751 0.784 0.72 ...
## $ Anu4 : num 0.927 0.958 0.816 0.878 0.799 ...
## $ Anu5 : num 0.94 0.988 0.836 0.998 0.832 ...
## $ Anu6 : num NA NA NA 1.04 NA ...
## $ Anu7 : num NA NA NA 1.08 NA ...
## $ Anu8 : num NA NA NA NA NA ...
## $ Anu9 : num NA NA NA NA NA ...
## $ RadCap: num 0.94 0.988 0.836 1.079 0.832 ...
## $ LenCap: int 347 292 348 374 375 355 195 289 388 423 ...
## $ WTg : int 658 415 557 669 716 719 118 479 986 1258 ...
## $ SEXCON: Factor w/ 5 levels "0","1","3","6",...: 5 3 3 5 3 5 5 3 3 5 ...
## $ Sex : Factor w/ 3 levels "0","1","2": 3 2 2 3 2 3 3 2 2 3 ...

```

```

### Removing Outliers
length(LMB$FID)

```

```
## [1] 132
```

```
LMB[LMB$FID==55,];FID55 <- as.numeric(rownames(LMB[LMB$FID==55,])) ### Data Entry Error
```

```

##      Month Day Year FID Site AgeCap  Anu1  Anu2  Anu3 Anu4 Anu5 Anu6 Anu7
## 55      6 30 2016 55    8      2 0.5196 0.7003 0.747  NA  NA  NA  NA
##      Anu8 Anu9 RadCap LenCap WTg SEXCON Sex
## 55      NA  NA  0.747   460 306      8    2

```

```

#LMB[LMB$FID==89,];FID89 <- as.numeric(rownames(LMB[LMB$FID==89,])) ### Unknown Sex > ezr model fitting
rm <- LMB[c(FID55),]; rm <- filterD(rm,!is.na(FID)) ### Create df containing removed fish
LMB <- LMB[-c(FID55),] %>%
  filterD(!is.na(FID)) ### Remove outliers
length(LMB$FID)

```

```
## [1] 131
```

```
#2-7-2018#write.csv(LMB,file="Data/Clean-Data/2016_largemouth-bass_clean.csv", row.names = FALSE)
```

Now I need to put the data into one observation per line (long) format.

```
LMBL <- gather(LMB, Agei, Radi, Anu1:Anu9) %>% arrange(FID, Agei)
```

Next I will clean up the Agei variable so that it contains only numbers, is numeric, has no NA values, and is always less than or equal to the age-at-capture.

```
str_sub(LMBL$Agei, start = 1, end = 3) <- ""
LMBL %<>% mutate(Agei = as.numeric(Agei)) %>% filterD(!is.na(Radi)) %>% filterD(Agei <=
  AgeCap)
```

Next I will perform my back-calculation using the Biological-Intercept model. Where Otoloth radius at the time of formation ( $O_0$ ) is 0.035 mm (Miller and Storck 1982) and the Length of the fish at time of otolith formation ( $L_0$ ) is 3.9 mm (Heang 1982).

```
LMBL$BI.len <- with(LMBL, LenCap + (Radi - RadCap) * (LenCap - 3.9) * (RadCap -
  0.035)^-1)
```

```
#2-7-2018#write.csv(LMBL, "Data/Clean-Data/2016_largemouth-bass_long-format.csv", row.names = FALSE)
```

I have commented out the final write.csv command because I do not want to accidentally remake my data file.